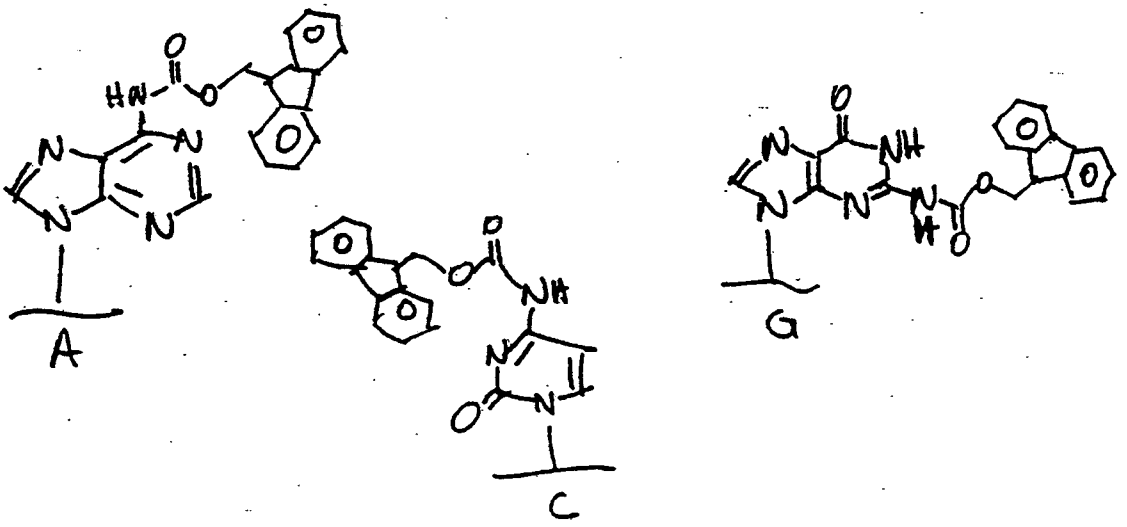
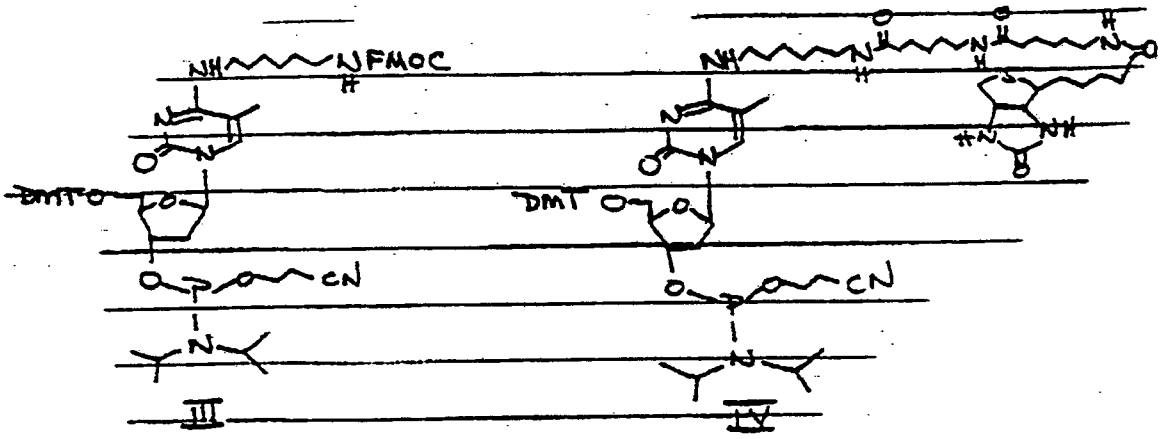
Another type of protecting group FMOC (9-fluorenyl methoxycarbonyl) is currently being used to protect the exocyclic amines of the three bases:



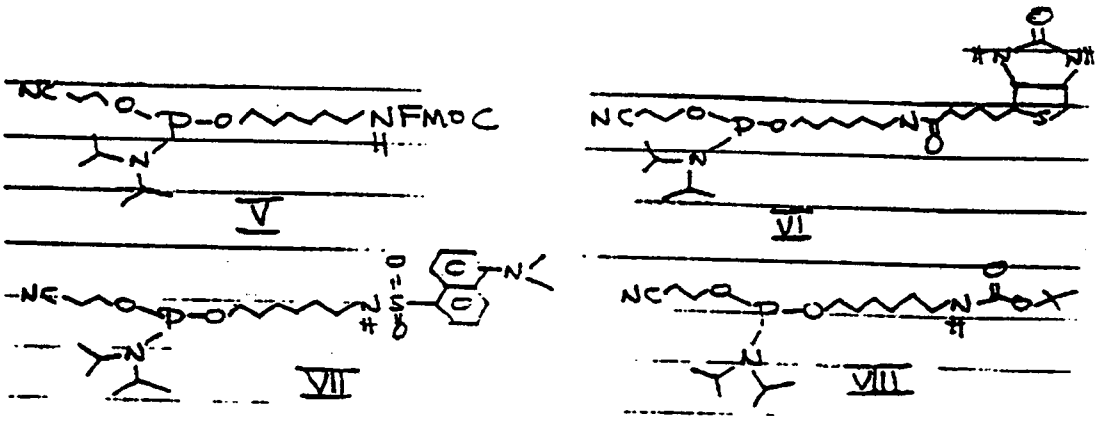
Adenine (A) Cytosine (C) Guanine (G)

The advantage of the FMOC group is that it is removed under mild conditions (dilute organic bases) and can be used for all three bases. The amide protecting groups require more harsh conditions to be removed (NH_3/MeOH with heat).

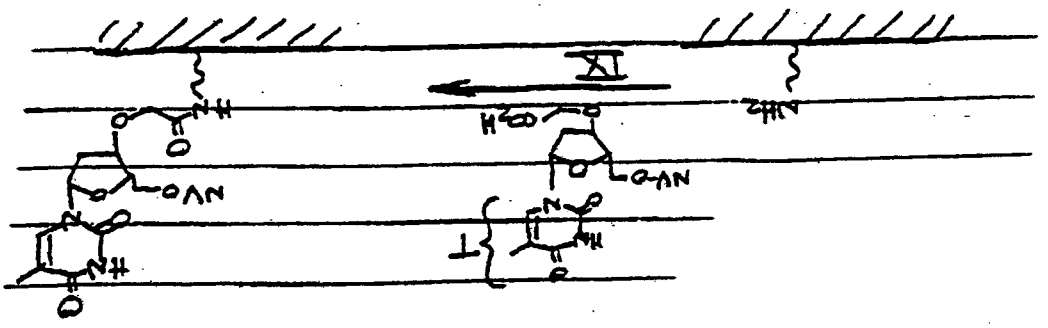
Nucleosides used as 5'-OH probes, useful in verifying correct VLSIPS synthetic function, have been the following:



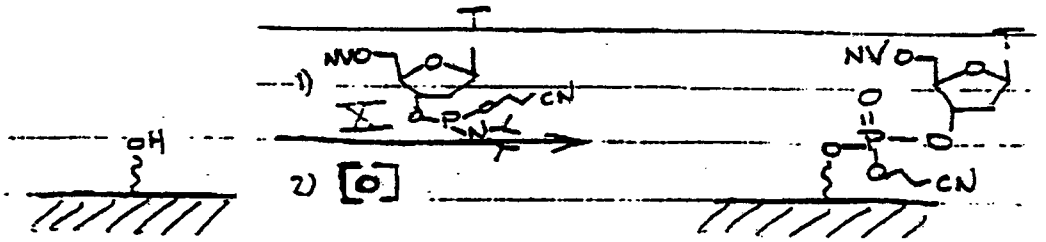
These compounds are used to detect where on a substrate photolysis has occurred by the attachment of either III or V to the newly generated 5'-OH. In the case of III, after the phosphate attachment is made, the substrate is treated with a dilute base to remove the FMOC group. The resulting amine can be reacted with FITC and the substrate examined by fluorescence microscopy. This indicates the proper generation of a 5'-OH. In the case of compound IV, after the phosphate attachment is made, the substrate is treated with FITC labeled streptavidin and the substrate again may be examined by fluorescence microscopy. Other probes, although not nucleoside based, have included the following:



The method of attachment of the first nucleoside to the surface of the substrate depends on the functionality of the groups at the substrate surface. If the surface is amine functionalized, an amide bond is made (see example below).



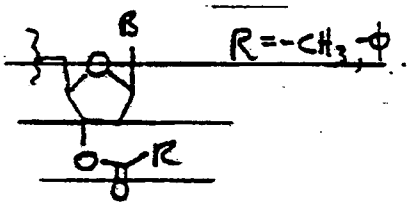
If the surface is hydroxy functionalized a phosphate bond is made (see example below)



In both cases, the thymidine example is illustrated, but any one of the four phosphoramidite activated nucleosides can be used in the first step.

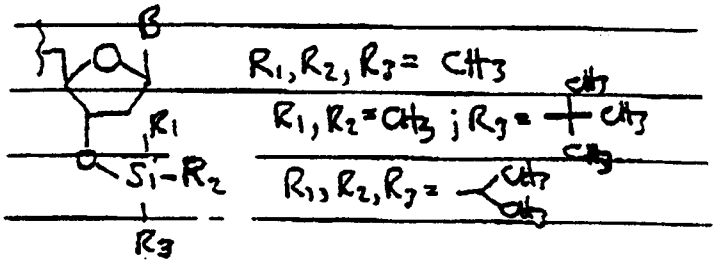
Photolysis of the photolabile group NV or NVOC on the 5' positions of the nucleosides is carried out at ~362 nm with an intensity of 14 mW/cm^2 for 10 minutes with the substrate side (side containing the photolabile group) immersed in dioxane. After the coupling of the next nucleoside is complete, the photolysis is repeated followed by another coupling until the desired oligomer is obtained.

One of the most common 3'-O-protecting group is the ester, in particular the acetate



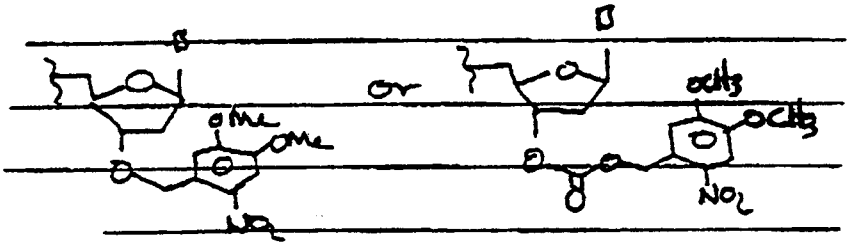
The groups can be removed by mild base treatment 0.1N NaOH/MeOH or K_2CO_3/H_2O/MeOH.

Another group used most often is the silyl ether.



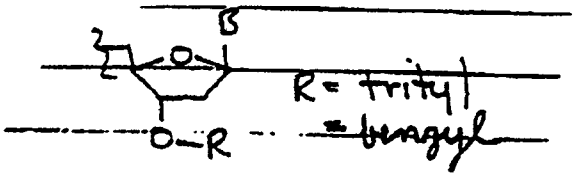
These groups can be removed by neutral conditions using 1 M tetra-n-butylammonium fluoride in THF or under acid conditions.

Related to photodeprotection, the nitroveratryl group could also be used to protect the 3'-position.



Here, light (photolysis) would be used to remove these protecting groups.

A variety of ethers can also be used in the protection of the 3'-O-position.



Removal of these groups usually involves acid or catalytic methods.

Note that corresponding linkages and photoblocked amino acids are described in detail in U.S.S.N. 07/624,120, which is hereby incorporated herein by reference.

Although the specificity of interactions at particular locations will usually be homogeneous due to a homogeneous polymer being synthesized at each defined location, for certain purposes, it may be useful to have mixed polymers with a commensurate mixed collection of interactions occurring at specific defined locations, or degeneracy reducing analogues, which have been discussed above and show broad specificity in binding. Then, a positive interaction signal may result from any of a number of sequences contained therein.

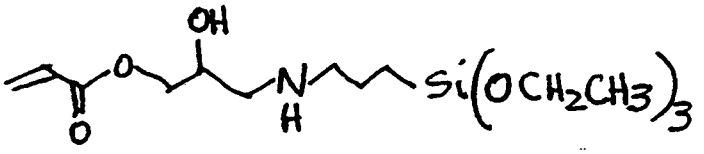
As an alternative method of generating a matrix pattern on a substrate, preformed polymers may be individually attached at particular sites on the substrate. This may be performed by individually attaching reagents one at a time to specific positions on the matrix, a process which may be automated. See, e.g., U.S.S.N. 07/435,316 (caged biotin parent), and U.S.S.N. 07/612,671 (caged biotin CIP). Another way of generating a positionally defined matrix pattern on a substrate is to have individually specific reagents which interact with each specific position on the substrate. For example, oligonucleotides may be synthesized at defined locations on the substrate. Then the substrate would have on its surface a plurality of regions having homogeneous oligonucleotides attached at each position.

In particular, at least four different substrate preparation procedures are available for treating a substrate surface. They are the standard VLSIPS method, polymeric substrates, Durapore™, and synthetic beads or fibers. The treatment labeled "standard VLSIPS" method is described in U.S.S.N. 07/624,120, and involves applying amino-propyltriethoxysilane to a glass surface.

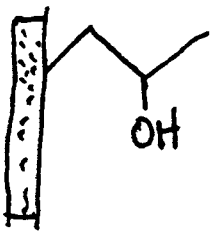
The polymeric substrate approach involves either of two ways of generating a polymeric substrate. The first uses a high concentration of aminopropyltriethoxysilane (2-20%) in an

aqueous ethanol solution (95%). This allows the silane compound to polymerize both in solution and on the substrate surface, which provides a high density of amines on the surface of the glass. This density is contrasted with the standard VLSIPS method. This polymeric method allows for the deposition on the substrate surface of a monolayer due to the anhydrous method used with the aforementioned silane.

The second polymeric method involves either the coating or covalent binding of an appropriate acrylic acid polymer onto the substrate surface. In particular, e.g., in DNA synthesis, a monomer such as a hydroxypropylacrylate is used to generate a high density of hydroxyl groups on the substrate surface, allowing for the formation of phosphate bonds. An example of such a compound is shown:



The method using a Durapore™ membrane (Millipore) consists of a polyvinylidine difluoride coating with crosslinked polyhydroxylpropyl acrylate [PVDF-HPA]:



Here the building up of, e.g., a DNA oligomer, can be started immediately since phosphate bonds to the surface can be accomplished in the first step with no need for modification. A nucleotide dimer (5'-C-T-3') has been successfully made on this substrate in our labs.

The fourth method utilizes synthetic beads or fibers. This would use another substrate, such as a teflon copolymer

graft bead or fiber, which is covalently coated with an organic layer (hydrophilic) terminating in hydroxyl sites (commercially available from Molecular Brosystems, Inc.) This would offer the same advantage as the Durapore™ membrane, allowing for immediate phosphate linkages, but would give additional contour by the 3-dimensional growth of oligomers.

A matrix pattern of new reagents may be targeted to each specific oligonucleotide position by attaching a complementary oligonucleotide to which the substrate bound form is complementary. For instance, a number of regions may have homogeneous oligonucleotides synthesized at various locations. Oligonucleotide sequences complementary to each of these can be individually generated and linked to a particular specific reagents. Often these specific reagents will be antibodies. As each of these is specific for finding its complementary oligonucleotide, each of the specific reagents will bind through the oligonucleotide to the appropriate matrix position. A single step having a combination of different specific reagents being attached specifically to a particular oligonucleotide will thereby bind to its complement at the defined matrix position. The oligonucleotides will typically then be covalently attached, using, e.g., an acridine dye, for photocrosslinking. Psoralen is a commonly used acridine dye for photocrosslinking purposes, see, e.g., Song et al. (1979) Photochem. Photobiol. 29:1177-1197; Cimino et al. (1985) Ann. Rev. Biochem. 54:1151-1193; Parsons (1980) Photochem. Photobiol. 32:813-821; and Dattagupta et al. (1985) U.S. Pat. No. 4,542,102, and (1987) U.S. Pat. No. 4,713,326; each of which is hereby incorporated herein by reference. This method allows a single attachment manipulation to attach all of the specific reagents to the matrix at defined positions and results in the specific reagents being homogeneously located at defined positions.

D. Surface Immobilization

1. caged biotin

An alternative method of attaching reagents in a positionally defined matrix pattern is to use a caged biotin

system. See U.S.S.N. 07/612,671 (caged biotin CIP), which is hereby incorporated herein by reference, for additional details on the chemistry and application of caged biotin embodiments. In short, the caged biotin has a photosensitive blocking moiety which prevents the combination of avidin to biotin. At positions where the photo-lithographic process has removed the blocking group, high affinity biotin sites are generated. Thus, by a sequential series of photolithographic deblocking steps interspersed with exposure of those regions to appropriate biotin containing reagents, only those locations where the deblocking takes place will form an avidin-biotin interaction. Because the avidin-biotin binding is very tight, this will usually be virtually irreversible binding.

2. crosslinked interactions

The surface immobilization may also take place by photocrosslinking of defined oligonucleotides linked to specific reagents. After hybridization of the complementary oligonucleotides, the oligonucleotides may be crosslinked by a reagent by psoralen or another similar type of acridine dye. Other useful crosslinking reagents are described in Dattagupta et al. (1985) U.S. Pat. No. 4,542,102, and (1987) U.S. Pat. No. 4,713,326.

In another embodiment, colony or phage plaque transfer of biological polymers may be transferred directly onto a silicon substrate. For example, a colony plate may be transferred onto a substrate having a generic oligonucleotide sequence which hybridizes to another generic complementary sequence contained on all of the vectors into which inserts are cloned. This will specifically only bind those molecules which are actually contained in the vectors containing the desired complementary sequence. This immobilization allows for producing a matrix onto which a sequence specific reagent can bind, or for other purposes. In a further embodiment, a plurality of different vectors each having a specific oligonucleotide attached to the vector may be specifically attached to particular regions on a matrix having a complementary oligonucleotide attached thereto.

VIII. HYBRIDIZATION/SPECIFIC INTERACTION

A. General

As discussed previously in the VLSIPS parent applications, the VLSIPS substrates may be used for screening for specific interactions with sequence specific targets or probes.

In addition, the availability of substrates having the entire repertoire of possible sequences of a defined length opens up the possibility of sequencing by hybridization. This sequence may be de novo determination of an unknown sequence, particularly of nucleic acid, verification of a sequence determined by another method, or an investigation of changes in a previously sequenced gene, locating and identifying specific changes. For example, often Maxam and Gilbert sequencing techniques are applied to sequences which have been determined by Sanger and Coulson. Each of those sequencing technologies have problems with resolving particular types of sequences. Sequencing by hybridization may serve as a third and independent method for verifying other sequencing techniques. See, e.g., (1988) Science 242:1245.

In addition, the ability to provide a large repertoire of particular sequences allows use of short subsequence and hybridization as a means to fingerprint a polynucleotide sample. For example, fingerprinting to a high degree of specificity of sequence matching may be used for identifying highly similar samples, e.g., those exhibiting high homology to the selected probes. This may provide a means for determining classifications of particular sequences. This should allow determination of whether particular genomes of bacteria, phage, or even higher cells might be related to one another.

In addition, fingerprinting may be used to identify an individual source of biological sample. See, e.g., Lander, E. (1989) Nature, 339:501-505, and references therein. For example, a DNA fingerprint may be used to determine whether a genetic sample arose from another individual. This would be particularly useful in various sorts of forensic tests to determine, e.g., paternity or sources of blood samples.

Significant detail on the particulars of genetic fingerprinting for identification purposes are described in, e.g., Morris et al. (1989) "Biostatistical evolution of evidence from continuous allele frequency distribution DNA probes in reference to disputed paternity of identity," J. Forensic Science 34:1311-1317; and Neufeld et al. (1990) Scientific American 262:46-53; each of which is hereby incorporated herein by reference.

In another embodiment, a fingerprinting-like procedure may be used for classifying cell types by analyzing a pattern of specific nucleic acids present in the cell, specifically RNA expression patterns. This may also be useful in defining the temporal stage of development of cells, e.g., stem cells or other cells which undergo temporal changes in development. For example, the stage of a cell, or group of cells, may be tested or defined by isolating a sample of mRNA from the population and testing to see what sequences are present in messenger populations. Direct samples, or amplified samples (e.g., by polymerase chain reaction), may be used. Where particular mRNA or other nucleic acid sequences may be characteristic of or shown to be characteristic of particular developmental stages, physiological states, or other conditions, this fingerprinting method may define them.

The present invention may also be used for mapping sequences within a larger segment. This may be performed by at least two methods, particularly in reference to nucleic acids. Often, enormous segments of DNA are subcloned into a large plurality of subsequences. Ordering these subsequences may be important in determining the overlaps of sequences upon nucleotide determinations. Mapping may be performed by immobilizing particularly large segments onto a matrix using the VLSIPS technology. Alternatively, sequences may be ordered by virtue of subsequences shared by overlapping segments. See, e.g., Craig et al. (1990) Nuc. Acids Res. 18:2653-2660; Michiels et al. (1987) CABIOS 3:203-210; and Olson et al. (1986) Proc. Natl. Acad. Sci. USA 83:7826-7830.

B. Important Parameters

The extent of specific interaction between reagents immobilized to the VLSIPS substrate and another sequence specific reagent may be modified by the conditions of the interaction. Sequencing embodiments typically require high fidelity hybridization and the ability to discriminate perfect matching from imperfect matching. Fingerprinting and mapping embodiments may be performed using less stringent conditions, or in some embodiments very highly stringent conditions, depending upon the circumstances.

In a nucleic acid hybridization embodiment, the specificity and kinetics of hybridization have been described in detail by, e.g., Wetmur and Davidson (1968) J. Mol. Biol., 31:349-370, Britten and Kohne (1968) Science 161:529-530, and Kanehisa, (1984) Nuc. Acids Res. 12:203-213, each of which is hereby incorporated herein by reference. Parameters which are well known to affect specificity and kinetics of reaction include salt conditions, ionic composition of the solvent, hybridization temperature, length of oligonucleotide matching sequences, guanine and cytosine (GC) content, presence of hybridization accelerators, pH, specific bases found in the matching sequences, solvent conditions, and addition of organic solvents.

In particular, the salt conditions required for driving highly mismatched sequences to completion typically include a high salt concentration. The typical salt used is sodium chloride (NaCl), however, other ionic salts may be utilized, e.g., KCl. Depending on the desired stringency hybridization, the salt concentration will often be less than about 3 molar, more often less than 2.5 molar, usually less than about 2 molar, and more usually less than about 1.5 molar. For applications directed towards higher stringency matching, the salt concentrations would typically be lower. Ordinary high stringency conditions will utilize salt concentration of less than about 1 molar, more often less then about 750 millimolar, usually less than about 500 millimolar, and may be as low as about 250 or 150 millimolar.

The kinetics of hybridization and the stringency of hybridization both depend upon the temperature at which the hybridization is performed and the temperature at which the washing steps are performed. Temperatures at which steps for low stringency hybridization are desired would typically be lower temperatures, e.g., ordinarily at least about 15°C, more ordinarily at least about 20°C, usually at least about 25°C, and more usually at least about 30°C. For those applications requiring high stringency hybridization, or fidelity of hybridization and sequence matching, temperatures at which hybridization and washing steps are performed would typically be high. For example, temperatures in excess of about 35°C would often be used, more often in excess of about 40°C, usually at least about 45°C, and occasionally even temperatures as high as about 50°C or 60°C or more. Of course, the hybridization of oligonucleotides may be disrupted by even higher temperatures. Thus, for stripping of targets from substrates, as discussed below, temperatures as high as 80°C, or even higher may be used.

The base composition of the specific oligonucleotides involved in hybridization affects the temperature of melting, and the stability of hybridization as discussed in the above references. However, the bias of GC rich sequences to hybridize faster and retain stability at higher temperatures can be compensated for by the inclusion in the hybridization incubation or wash steps of various buffers. Sample buffers which accomplish this result include the triethly-and trimethyl ammonium buffers. See, e.g., Wood et al. (1987) Proc. Natl. Acad. Sci. USA, 82:1585-1588, and Khrapko, K. et al. (1989) FEBS Letters 256:118-122.

The rate of hybridization can also be affected by the inclusion of particular hybridization accelerators. These hybridization accelerators include the volume exclusion agents characterized by dextran sulfate, or polyethylene glycol (PEG). Dextran sulfate is typically included at a concentration of between 1% and 40% by weight. The actual concentration selected depends upon the application, but typically a faster hybridization is desired in which the concentration is

optimized for the system in question. Dextran sulfate is often included at a concentration of between 0.5% and 2% by weight or dextran sulfate at a concentration between about 0.5% and 5%. Alternatively, proteins which accelerate hybridization may be added, e.g., the recA protein found in E. coli) or other homologous proteins.

Of course, the specific hybridization conditions will be selected to correspond to a discriminatory condition which provides a positive signal where desired but fails to show a positive signal at affinities where interaction is not desired. This may be determined by a number of titration steps or with a number of controls which will be run during the hybridization and/or washing steps to determine at what point the hybridization conditions have reached the stage of desired specificity.

IX. DETECTION METHODS

Methods for detection depend upon the label selected. The criteria for selecting an appropriate label are discussed below, however, a fluorescent label is preferred because of its extreme sensitivity and simplicity. Standard labeling procedures are used to determine the positions where interactions between a sequence and a reagent take place. For example, if a target sequence is labeled and exposed to a matrix of different probes, only those locations where probes do interact with the target will exhibit any signal. Alternatively, other methods may be used to scan the matrix to determine where interaction takes place. Of course, the spectrum of interactions may be determined in a temporal manner by repeated scans of interactions which occur at each of a multiplicity of conditions. However, instead of testing each individual interaction separately, a multiplicity of sequence interactions may be simultaneously determined on a matrix.

A. Labeling Techniques

The target polynucleotide may be labeled by any of a number of convenient detectable markers. A fluorescent label is preferred because it provides a very strong signal with low

background. It is also optically detectable at high resolution and sensitivity through a quick scanning procedure. Other potential labeling moieties include, radioisotopes, chemiluminescent compounds, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, and linked enzymes.

Another method for labeling does not require incorporation of a labeling moiety. The target may be exposed to the probes, and a double strand hybrid is formed at those positions only. Addition of a double strand specific reagent will detect where hybridization takes place. An intercalative dye such as ethidium bromide may be used as long as the probes themselves do not fold back on themselves to a significant extent forming hairpin loops. See, e.g., Sheldon et al. (1986) U.S. Pat. No. 4,582,789. However, the length of the hairpin loops in short oligonucleotide probes would typically be insufficient to form a stable duplex.

In another embodiment, different targets may be simultaneously sequenced where each target has a different label. For instance, one target could have a green fluorescent label and a second target could have a red fluorescent label. The scanning step will distinguish sites of binding of the red label from those binding the green fluorescent label. Each sequence can be analyzed independently from one another.

Suitable chromogens will include molecules and compounds which absorb light in a distinctive range of wavelengths so that a color may be observed, or emit light when irradiated with radiation of a particular wave length or wave length range, e.g., fluorescers.

A wide variety of suitable dyes are available, being primary chosen to provide an intense color with minimal absorption by their surroundings. Illustrative dye types include quinoline dyes, triarylmethane dyes, acridine dyes, alizarine dyes, phthaleins, insect dyes, azo dyes, anthraquinoid dyes, cyanine dyes, phenazathionium dyes, and phenazoxonium dyes.

A wide variety of fluorescers may be employed either by themselves or in conjunction with quencher molecules.

Fluorescers of interest fall into a variety of categories having certain primary functionalities. These primary functionalities include 1- and 2-aminonaphthalene, p,p'-diaminostilbenes, pyrenes, quaternary phenanthridine salts, 9-aminoacridines, p,p'-diaminobenzophenone imines, anthracenes, oxacarbocyanine, merocyanine, 3-aminoequilenin, perylene, bis-benzoxazole, bis-p-oxazolyl benzene, 1,2-benzophenazin, retinol, bis-3-aminopyridinium salts, hellebrigenin, tetracycline, sterophenol, benzimidzaolylphenylamine, 2-oxo-3-chromen, indole, xanthen, 7-hydroxycoumarin, phenoxazine, salicylate, strophanthidin, porphyrins, triarylmethanes and flavin. Individual fluorescent compounds which have functionalities for linking or which can be modified to incorporate such functionalities include, e.g., dansyl chloride; fluoresceins such as 3,6-dihydroxy-9-phenylxanthhydrol; rhodamineisothiocyanate; N-phenyl 1-amino-8-sulfonatonaphthalene; N-phenyl 2-amino-6-sulfonatonaphthalene; 4-acetamido-4-isothiocyanato-stilbene-2,2'-disulfonic acid; pyrene-3-sulfonic acid; 2-toluidinonaphthalene-6-sulfonate; N-phenyl, N-methyl 2-aminoaphthalene-6-sulfonate; ethidium bromide; stebrine; auromine-0,2-(9'-anthroyl)palmitate; dansyl phosphatidylethanolamine; N,N'-dioctadecyl oxacarbocyanine; N,N'-dihexyl oxacarbocyanine; merocyanine, 4-(3'pyrenyl)butyrate; d-3-aminodesoxy-equilenin; 12-(9'-anthroyl)stearate; 2-methylanthracene; 9-vinylanthracene; 2,2'-(vinylene-p-phenylene)bisbenzoxazole; p-bis[2-(4-methyl-5-phenyl-oxazolyl)]benzene; 6-dimethylamino-1,2-benzophenazin; retinol; bis(3'-aminopyridinium) 1,10-decandiyl diiodide; sulfonaphthylhydrazone of hellibrienin; chlorotetracycline; N-(7-dimethylamino-4-methyl-2-oxo-3-chromenyl)maleimide; N-[p-(2-benzimidazolyl)-phenyl]maleimide; N-(4-fluoranthyl)maleimide; bis(homovanillic acid); resazarin; 4-chloro-7-nitro-2,1,3-benzooxadiazole; merocyanine 540; resorufin; rose bengal; and 2,4-diphenyl-3(2H)-furanone.

Desirably, fluorescers should absorb light above about 300 nm, preferably about 350 nm, and more preferably above about 400 nm, usually emitting at wavelengths greater than about 10 nm higher than the wavelength of the light

absorbed. It should be noted that the absorption and emission characteristics of the bound dye may differ from the unbound dye. Therefore, when referring to the various wavelength ranges and characteristics of the dyes, it is intended to indicate the dyes as employed and not the dye which is unconjugated and characterized in an arbitrary solvent.

Fluorescers are generally preferred because by irradiating a fluorescer with light, one can obtain a plurality of emissions. Thus, a single label can provide for a plurality of measurable events.

Detectable signal may also be provided by chemiluminescent and bioluminescent sources. Chemiluminescent sources include a compound which becomes electronically excited by a chemical reaction and may then emit light which serves as the detectible signal or donates energy to a fluorescent acceptor. A diverse number of families of compounds have been found to provide chemiluminescence under a variety of conditions. One family of compounds is 2,3-dihydro-1,-4-phthalazinedione. The most popular compound is luminol, which is the 5-amino compound. Other members of the family include the 5-amino-6,7,8-trimethoxy- and the dimethylamino[ca]benz analog. These compounds can be made to luminesce with alkaline hydrogen peroxide or calcium hypochlorite and base. Another family of compounds is the 2,4,5-triphenylimidazoles, with lophine as the common name for the parent product. Chemiluminescent analogs include para-dimethylamino and -methoxy substituents. Chemiluminescence may also be obtained with oxalates, usually oxalyl active esters, e.g., p-nitrophenyl and a peroxide, e.g., hydrogen peroxide, under basic conditions. Alternatively, luciferins may be used in conjunction with luciferase or lucigenins to provide bioluminescence.

Spin labels are provided by reporter molecules with an unpaired electron spin which can be detected by electron spin resonance (ESR) spectroscopy. Exemplary spin labels include organic free radicals, transitional metal complexes, particularly vanadium, copper, iron, and manganese, and the like. Exemplary spin labels include nitroxide free radicals.

B. Scanning System

With the automated detection apparatus, the correlation of specific positional labeling is converted to the presence on the target of sequences for which the reagents have specificity of interaction. Thus, the positional information is directly converted to a database indicating what sequence interactions have occurred. For example, in a nucleic acid hybridization application, the sequences which have interacted between the substrate matrix and the target molecule can be directly listed from the positional information. The detection system used is described in PCT publication no. WO90/15070; and U.S.S.N. 07/624,120. Although the detection described therein is a fluorescence detector, the detector may be replaced by a spectroscopic or other detector. The scanning system may make use of a moving detector relative to a fixed substrate, a fixed detector with a moving substrate, or a combination. Alternatively, mirrors or other apparatus can be used to transfer the signal directly to the detector. See, e.g, U.S.S.N. 07/624,120, which is hereby incorporated herein by reference.

The detection method will typically also incorporate some signal processing to determine whether the signal at a particular matrix position is a true positive or may be a spurious signal. For example, a signal from a region which has actual positive signal may tend to spread over and provide a positive signal in an adjacent region which actually should not have one. This may occur, e.g., where the scanning system is not properly discriminating with sufficiently high resolution in its pixel density to separate the two regions. Thus, the signal over the spatial region may be evaluated pixel by pixel to determine the locations and the actual extent of positive signal. A true positive signal should, in theory, show a uniform signal at each pixel location. Thus, processing by plotting number of pixels with actual signal intensity should have a clearly uniform signal intensity. Regions where the signal intensities show a fairly wide dispersion, may be

particularly suspect and the scanning system may be programmed to more carefully scan those positions.

In another embodiment, as the sequence of a target is determined at a particular location, the overlap for the sequence would necessarily have a known sequence. Thus, the system can compare the possibilities for the next adjacent position and look at these in comparison with each other. Typically, only one of the possible adjacent sequences should give a positive signal and the system might be programmed to compare each of these possibilities and select that one which gives a strong positive. In this way, the system can also simultaneously provide some means of measuring the reliability of the determination by indicating what the average signal to background ratio actually is.

More sophisticated signal processing techniques can be applied to the initial determination of whether a positive signal exists or not. See, e.g., U.S.S.N. 07/624,120.

From a listing of those sequences which interact, data analysis may be performed on a series of sequences. For example, in a nucleic acid sequence application, each of the sequences may be analyzed for their overlap regions and the original target sequence may be reconstructed from the collection of specific subsequences obtained therein. Other sorts of analyses for different applications may also be performed, and because the scanning system directly interfaces with a computer the information need not be transferred manually. This provides for the ability to handle large amounts of data with very little human intervention. This, of course, provides significant advantages over manual manipulations. Increased throughput and reproducibility is thereby provided by the automation of vast majority of steps in any of these applications.

DATA ANALYSIS

A. General

Data analysis will typically involve aligning the proper sequences with their overlaps to determine the target sequence. Although the target "sequence" may not specifically

correspond to any specific molecule, especially where the target sequence is broken and fragmented up in the sequencing process, the sequence corresponds to a contiguous sequence of the subfragments.

The data analysis can be performed by a computer using an appropriate program. See, e.g., Drmanac, R. et al. (1989) Genomics 4:114-128; and a commercially available analysis program available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia. Although the specific manipulations necessary to reassemble the target sequence from fragments may take many forms, one embodiment uses a sorting program to sort all of the subsequences using a defined hierarchy. The hierarchy need not necessarily correspond to any physical hierarchy, but provides a means to determine, in order, which subfragments have actually been found in the target sequence. In this manner, overlaps can be checked and found directly rather than having to search throughout the entire set after each selection process. For example, where the oligonucleotide probes are 10-mers, the first 9 positions can be sorted. A particular subsequence can be selected as in the examples, to determine where the process starts. As analogous to the theoretical example provided above, the sorting procedure provides the ability to immediately find the position of the subsequence which contains the first 9 positions and can compare whether there exists more than 1 subsequence during the first 9 positions. In fact, the computer can easily generate all of the possible target sequences which contain given combination of subsequences. Typically there will be only one, but in various situations, there will be more.

An exemplary flow chart for a sequencing program is provided in Figure 4. In general terms, the program provides for automated scanning of the substrate to determine the positions of probe and target interaction. Simple processing of the intensity of the signal may be incorporated to filter out clearly spurious signals. The positions with positive interaction are correlated with the sequence specificity of specific matrix positions, to generate the set of matching

subsequences. This information is further correlated with other target sequence information, e.g., restriction fragment analysis. The sequences are then aligned using overlap data, thereby leading to possible corresponding target sequences which will, optimally, correspond to a single target sequence.

B. Hardware

A variety of computer systems may be used to run a sequencing program. The program may be written to provide both the detecting and scanning steps together and will typically be dedicated to a particular scanning apparatus. However, the components and functional steps may be separated and the scanning system may provide an output, e.g., through tape or an electronic connection into a separate computer which separately runs the sequencing analysis program. The computer may be any of a number of machines provided by standard computer manufacturers, e.g., IBM compatible machines, Apple™ machines, VAX machines, and others, which may often use a UNIX™ operating system. Alternatively, custom computing architectures may be employed, these architectures may include neural network methods implemented in hardware and/or software. Of course, the hardware used to run the analysis program will typically determine what programming language would be used.

C. Software

Software would be readily developed by a person of ordinary skill in the programming art, following the flow chart provided, or based upon the input provided and the desired result.

Of course, an exemplary embodiment is a polynucleotide sequence system. However, the theoretical and mathematical manipulations necessary for data analysis of other linear molecules are conceptually similar.

XI. SUBSTRATE REUSE

Where a substrate is made with specific reagents that are relatively insensitive to the handling and processing steps involved in a single cycle of use, the substrate may often be

reused. The target molecules are usually stripped off of the solid phase specific recognition molecules. Of course, it is preferred that the manipulations and conditions be selected as to be mild and to not affect the substrate. For example, if a substrate is acid labile, a neutral pH would be preferred in all handling steps. Similar sensitivities would be carefully respected where recycling is desired.

A. Removal of Label

Typically for a recycling, the previously attached specific interaction would be disrupted and removed. This will typically involve exposing the substrate to conditions under which the interaction between probe and target is disrupted. Alternatively, it may be exposed to conditions where the target is destroyed. For example, where the probes are oligonucleotides and the target is a polynucleotide, a heating and low salt wash will often be sufficient to disrupt the interactions. Additional reagents may be added such as detergents, and organic or inorganic solvents which disrupt the interaction between the specific reagents and target.

B. Storage and Preservation

As indicated above, the matrix will typically be maintained under conditions where the matrix itself and the linkages and specific reagents are preserved. Various specific preservatives may be added which prevent degradation. For example, if the reagents are acid or base labile, a neutral pH buffer will typically be added. It is also desired to avoid destruction of the matrix by growth of organisms which may destroy organic reagents attached thereto. For this reason, a preservative such as cyanide or azide may be added. However, the chemical preservative should also be selected to preserve the chemical nature of the linkages and other components of the substrate. Typically, a detergent may also be included.

C. Processes to Avoid Degradation of Oligomers

In particular, a substrate comprising a large number of oligomers will be treated in a fashion which is known to

maintain the quality and integrity of oligonucleotides. These include storing the substrate in a carefully controlled environment under conditions of lower temperature, cation depletion (EDTA and EGTA), sterile conditions, and inert argon or nitrogen atmosphere.

XII. INTEGRATED SEQUENCING STRATEGY

A. Initial Mapping Strategy

As indicated above, although the VLSIPS may be applied to sequencing embodiments, it is often useful to integrate other concepts to simply the sequencing. For example, nucleic acids may be easily sequenced by careful selection of the vectors and hosts used for amplifying and generating the specific target sequences. For example, it may be desired to use specific vectors which have been designed to interact most efficiently with the VLSIPS substrate. This is also important in fingerprinting and mapping strategies. For example, vectors may be carefully selected having particular complementary sequences which are designed to attach to a genetic or specific oligomer on the substrate. This is also applicable to situations where it is desired to target particular sequences to specific locations on the matrix.

In one embodiment, unnatural oligomers may be used to target natural probes to specific locations on the VLSIPS substrate. In addition, particular probes may be generated for the mapping embodiment which are designed to have specific combinations of characteristics. For example, the construction of a mapping substrate may depend upon use of another automated apparatus which takes clones isolated from a chromosome walk and attaches them individually or in bulk to the VLSIPS substrate.

In another embodiment, a variety of specific vectors having known and particular "targeting" sequences adjacent the cloning sites may be individually used to clone a selected probe, and the isolated probe will then be targetable to a site on the VLSIPS substrate with a sequence complementary to the "target" sequence.

B. Selection of Smaller Clones

In the fingerprinting and mapping embodiments, the selection of probes may be very important. Significant mathematical analysis may be applied to determine which specific sequences should be used as those probes. Of course, for fingerprinting use, sequences that show significant heterogeneity across the human population would be preferred. Selection of the specific sequences which would most favorably be utilized will tend to be single copy sequences within the genome, and more specifically single copy sequences that have low cross-hybridization potential to other sequences in the genome (i.e., not members of a closely-related multigene family).

Various hybridization selection procedures may be applied to select sequences which tend not to be repeated within a genome, and thus would tend to be conserved across individuals. For example, hybridization selections may be made for non-repetitive and single copy sequences. See, e.g., Britten and Kohne (1968) "Repeated Sequences in DNA," Science 161:529-540. On the other hand, it may be desired under certain circumstances to use repeated sequences. For example, where a fingerprint may be used to identify or distinguish different species, or where repetitive sequences may be diagnostic of specific species, repetitive sequences may be desired for inclusion in the fingerprinting probes. In either case, the sequencing capability will greatly assist in the selection of appropriate sequences to be used as probes.

Also as indicated above, various means for constructing an appropriate substrate may involve either mechanical or automated procedures. The standard VLSIPS automated procedure involves synthesizing oligonucleotides or short polymers directly on the substrate. In various other embodiments, it is possible to attach separately synthesized reagents onto the matrix in an ordered array. Other circumstances may lend themselves to transfer a pattern from a petri plate onto a solid substrate. Also, there are methods for site specifically directing collections of reagents to

specific locations using unnatural nucleotides or equivalent sorts of targeting molecules.

While a brute force manual transfer process may be utilized sequentially attaching various samples to successive positions, instrumentation for automating such procedures may also be devised. The automated system for performing such would preferably be relatively easily designed and conceptually easily understood.

XIII. COMMERCIAL APPLICATIONS

A. Sequencing

As indicated above, sequencing may be performed either de novo or as a verification of another sequencing method. The present hybridization technology provides the ability to sequence nucleic acids and polynucleotides de novo, or as a means to verify either the Maxam and Gilbert chemical sequencing technique or Sanger and Coulson dideoxy- sequencing techniques. The hybridization method is useful to verify sequencing determined by any other sequencing technique and to closely compare two similar sequences, e.g., to identify and locate sequence differences.

Of course, sequencing of can be very important in many different sorts of environments. For example, it will be useful in determining the genetic sequence of particular markers in various individuals. In addition, polymers may be used as markers or for information containing molecules to encode information. For example, a short polynucleotide sequence may be included in large bulk production samples indicating the manufacturer, date, and location of manufacture of a product. For example, various drugs may be encoded with this information with a small number of molecules in a batch. For example, a pill may have somewhere from 10 to 100 to 1,000 or more very short and small molecules encoding this information. When necessary, this information may be decoded from a sample of the material using a polymerase chain reaction (PCR) or other amplification method. This encoding system may be used to provide the origin of large bulky samples without significantly affecting the properties of those samples. For

example, chemical samples may also be encoded by this method thereby providing means for identifying the source and manufacturing details of lots. The origin of bulk hydrocarbon samples may be encoded. Production lots of organic compounds such as benzene or plastics may be encoded with a short molecule polymer. Food stuffs may also be encoded using similar marking molecules. Even toxic waste samples can be encoded determining the source or origin. In this way, proper disposal can be traced or more easily enforced.

Similar sorts of encoding may be provided by fingerprinting-type analysis. Whether the resolution is absolute or less so, the concept of coding information on molecules such as nucleic acids, which can be amplified and later decoded, may be a very useful and important application.

This technology also provides the ability to include markers for origins of biological materials. For example, a patented animal line may be transformed with a particular unnatural sequence which can be traced back to its origin. With a selection of multiple markers, the likelihood could be negligible that a combination of markers would have independently arisen from a source other than the patented or specifically protected source. This technique may provide a means for tracing the actual origin of particular biological materials. Bacteria, plants, and animals will be subject to marking by such encoding sequences.

B. Fingerprinting

As indicated above, fingerprinting technology may also be used for data encryption. Moreover, fingerprinting allows for significant identification of particular individuals. Where the fingerprinting technology is standardized, and used for identification of large numbers of people, related equipment and peripheral processing will be developed to accompany the underlying technology. For example, specific equipment may be developed for automatically taking a biological sample and generating or amplifying the information molecules within the sample to be used in fingerprinting analysis. Moreover, the fingerprinting substrate may be mass

produced using particular types of automatic equipment. Synthetic equipment may produce the entire matrix simultaneously by stepwise synthetic methods as provided by the VLSIPS technology. The attachment of specific probes onto a substrate may also be automated, e.g., making use of the caged biotin technology. See, e.g., U.S.S.N. 07/612,671 (caged biotin CIP).

In addition, peripheral processing may be important and may be dedicated to this specific application. Thus, automated equipment for producing the substrates may be designed, or particular systems which take in a biological sample and output either a computer readout or an encoded instrument, e.g., a card or document which indicates the information and can provide that information to others. An identification having a short magnetic strip with a few million bits may be used to provide individual identification and important medical information useful in a medical emergency.

In fact, data banks may be set up to correlate all of this information of fingerprinting with medical information. This may allow for the determination of correlations between various medical problems and specific DNA sequences. By collating large populations of medical records with genetic information, genetic propensities and genetic susceptibilities to particular medical conditions may be developed. Moreover, with standardization of substrates, the micro encoding data may be also standardized to reproduce the information from a centralized data bank or on an encoding device carried on an individual person. On the other hand, if the fingerprinting procedure is sufficiently quick and routine, every hospital may routinely perform a fingerprinting operation and from that determine many important medical parameters for an individual.

In particular industries, the VLSIPS sequencing, fingerprinting, or mapping technology will be particularly appropriate. As mentioned above, agricultural livestock suppliers may be able to encode and determine whether their particular strains are being used by others. By incorporating particular markers into their genetic stocks, the markers will indicate origin of genetic material. This is applicable to

seed producers, livestock producers, and other suppliers of medical or agricultural biological materials.

This may also be useful in identifying individual animals or plants. For example, these markers may be useful in determining whether certain fish return to their original breeding grounds, whether sea turtles always return to their original birthplaces, or to determine the migration patterns and viability of populations of particular endangered species. It would also provide means for tracking the sources of particular animal products. For example, it might be useful for determining the origins of controlled animal substances such as elephant ivory or particular bird populations whose importation or exportation is controlled.

As indicated above, polymers may be used to encode important information on source and batch and supplier. This is described in greater detail, e.g., "Applications of PCR to industrial problems," (1990) in Chemical and Engineering News 68:145, which is hereby incorporated herein by reference. In fact, the synthetic method can be applied to the storage of enormous amounts of information. Small substrates may encode enormous amounts of information, and its recovery will make use of the inherent replication capacity. For example, on regions of 10 μm x 10 μm, 1 cm^2 has 10^6 regions. An theory, the entire human genome could be attached in 1000 nucleotide segments on a 3 cm^2 surface. Genomes of endangered species may be stored on these substrates.

Fingerprinting may also be used for genetic tracing or for identifying individuals for forensic science purposes. See, e.g., Morris, J. et al. (1989) "Biostatistical Evaluation of Evidence From Continuous Allele Frequency Distribution DNA Probes in Reference to Disputed Paternity and Identity," J. Forensic Science 34:1311-1317, and references provided therein; each of which is hereby incorporated herein by reference.

In addition, the high resolution fingerprinting allows the distinguishability to high resolution of particular samples. As indicated above, new cell classifications may be defined based on combinations of a large number of properties. Similar applications will be found in distinguishing different

species of animals or plants. In fact, microbial identification may become dependent on characterization of the genetic content. Tumors or other cells exhibiting abnormal physiology will be detectable by use of the present invention. Also, knowing the genetic fingerprint of a microorganism may provide very useful information on how to treat an infection by such organism.

Modifications of the fingerprint embodiments may be used to diagnose the condition of the organism. For example, a blood sample is presently used for diagnosing any of a number of different physiological conditions. A multi-dimensional fingerprinting method made available by the present invention could become a routine means for diagnosing an enormous number of physiological features simultaneously. This may revolutionize the practice of medicine in providing information on an enormous number of parameters together at one time. In another way, the genetic predisposition may also revolutionize the practice of medicine providing a physician with the ability to predict the likelihood of particular medical conditions arising at any particular moment. It also provides the ability to apply preventative medicine.

Also available are kits with the reagents useful for performing sequencing, fingerprinting, and mapping procedures. The kits will have various compartments with the desired necessary reagents, e.g., substrate, labeling reagents for target samples, buffers, and other useful accompanying products.

C. Mapping

The present invention also provides the means for mapping sequences within enormous stretches of sequence. For example, nucleotide sequences may be mapped within enormous chromosome size sequence maps. For example, it would be possible to map a chromosomal location within the chromosome which contains hundreds of millions of nucleotide base pairs. In addition, the mapping and fingerprinting embodiments allow for testing of chromosomal translocations, one of the standard problems for which amniocentesis is performed.

The present invention will be better understood by reference to the following illustrative examples. The following examples are offered by way of illustration and not by way of limitation.

Relevant applications whose techniques are incorporated herein by reference are PCT publication no. WO90/15070, published December 13, 1990; PCT publication no. WO91/07087, published May 30, 1991; U.S.S.N. 07/624,120, filed December 6, 1990; and U.S.S.N. 07/626,730, filed December 6, 1990.

Also, additional relevant techniques are described, e.g., in Sambrook, J., et al. (1989) Molecular Cloning: a Laboratory Manual, 2d Ed., vols 1-3, Cold Spring Harbor Press, New York; Greenstein and Winitz (1961) Chemistry of the Amino Acids, Wiley and Sons, New York; Bodzansky, M. (1988) Peptide Chemistry: a Practical Textbook, Springer-Verlag, New York; Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Press, New York; Glover, D. (ed.) (1987) DNA Cloning: A Practical Approach, vols 1-3, IRL Press, Oxford; Bishop and Rawlings (1987) Nucleic Acid and Protein Sequence Analysis: A Practical Approach, IRL Press, Oxford; Hames and Higgins (1985) Nucleic Acid Hybridisation: A Practical Approach, IRL Press, Oxford; Wu et al. (1989) Recombinant DNA Methodology, Academic Press, San Diego; Goding (1986) Monoclonal Antibodies: Principles and Practice, (2d ed.), Academic Press, San Diego; Finegold and Barron (1986) Bailey and Scott's Diagnostic Microbiology, (7th ed.), Mosby Co., St. Louis; Collins et al. (1989) Microbiological Methods, (6th ed.), Butterworth, London; Chaplin and Kennedy (1986) Carbohydrate Analysis: A Practical Approach, IRL Press, Oxford; Van Dyke (ed.) (1985) Bioluminescence and Chemiluminescence: Instruments and Applications, vol 1, CRC Press, Boca Rotan; and Ausubel et al. (ed.) (1990) Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York; each of which is hereby incorporated herein by reference.

EXAMPLES

The following examples are provided to illustrate the efficacy of the inventions herein. All operations were conducted at about ambient temperatures and pressures unless indicated to the contrary.

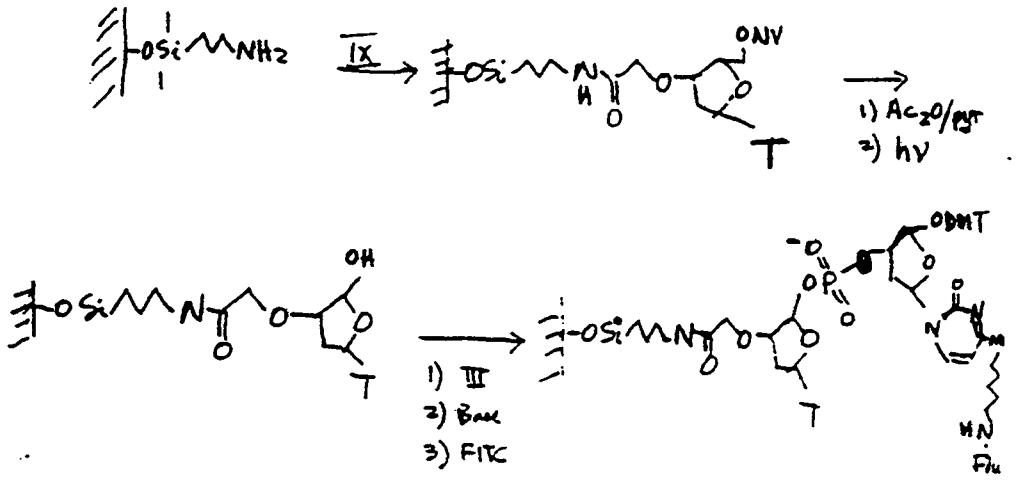
POLYNUCLEOTIDE SEQUENCING

1. HPLC of the photolysis of 5'-O-nitroveratryl-thymidine.

In order to determine the time for photolysis of 5'-O-nitrovertryl thymidine to thymidine a 100 μM solution of NV-Thym-OH (5'-O-nitrovertryl thymidine) in dioxane was made and ~200 μl aliquots were irradiated (in a quartz cuvette 1 cm x 2 mm) at 362.3 nm for 20 sec, 40 sec, 60 sec, 2 min, 5 min, 10 min, 15 min, and 20 min. The resulting irradiated mixtures were then analyzed by HPLC using a Varian MicroPak SP column (C_{18} analytical) at a flow rate of 1 ml/min and a solvent system of 40% CH_3CN and 60% water. Thymidine has a retention time of 1.2 min and NVO-Thym-OH has a retention time of 2.1 min. It was seen that after 10 min of exposure the deprotection was complete.

2. Preparation and Detection of Thymidine-Cytidine dimer (FITC)

The reaction is illustrated:



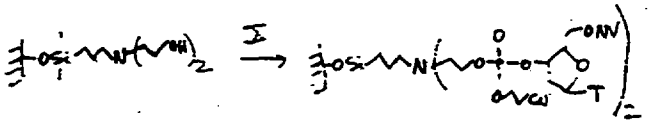
To an aminopropylated glass slide (standard VLSIPS) was added a mixture of the following:

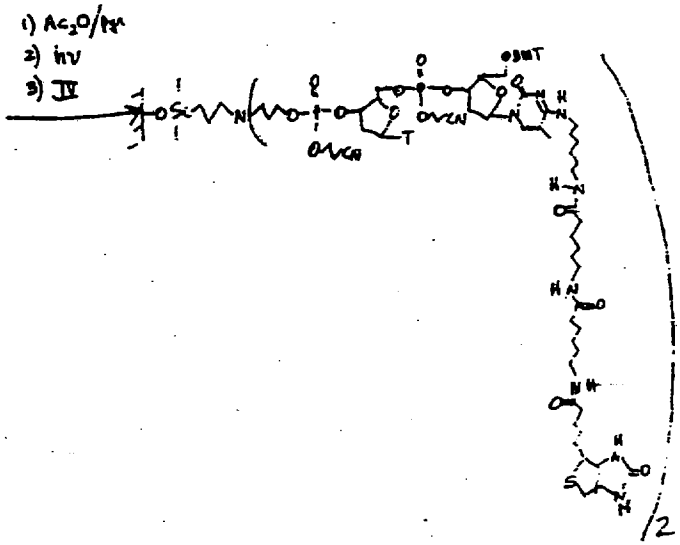
12.2 mg of NVO-Thym-CO_2H (IX)
3.4 mg of HOBT (N-hydroxybenztriazal)
8.8 µl DIEA (Diisopropylethylamine)
11.1 mg BOP reagent
2.5 ml DMF

After 2 h coupling time (standard VLSIPS) the plate was washed, acetylated with acetic anhydride/pyridine, washed, dried, and photolyzed in dioxane at 362 nm at 14 mW/cm^2 for 10 min using a 500 µm checkerboard mask. The slide was then taken and treated with a mixture of the following:

107 mg of FMOC-amine modified C (III)
21 mg of tetrazole
1 ml anhydrous CH_3CN

After being treated for approximately 8 min, the slide was washed off with CH_3CN, dried, and oxidized with I_2/H_2O/THF/lutidine for 1 min. The slide was again washed, dried, and treated for 30 min with a 20% solution of DBU in DMF. After thorough rinsing of the slide, it was next exposed to a FITC solution (1mM fluorescein isothiocyanate [FITC] in DMF) for 50 min, then washed, dried, and examined by fluorescence microscopy. This reaction is illustrated:





3. Preparation and Detection of Thymidine-Cytidine dimer (Biotin)

An aminopropyl glass slide, was soaked in a solution of ethylene oxide (20% in DMF) to generate a hydroxylated surface. The slide was added a mixture of the following:

32 mg of NVO-T-OCED (X)

11 mg of tetrazole

0.5 ml of anhydrous CH_3CN

After 8 min the plate was then rinsed with acetonitrile, then oxidized with I_2/H_2O/THF/lutidine for 1 min, washed and dried. The slide was then exposed to a 1:3 mixture of acetic anhydride:pyridine for 1 h, then washed and dried. The substrate was a then photolyzed in dioxane at 362 nm at 14 mW/cm^2 for 10 min using a 500μm checkerboard mask, dried, and then treated with a mixture of the following:

65 mg of biotin modified C (IV)

11 mg of tetrazole

0.5 ml anhydrous CH_3CN

After 8 min the slide was washed with CH_3CN then oxidized with I_2/H_2O/THF/lutidine for 1 min, washed, and then dried. The slide was then soaked for 30 min in a PBS/0.05% Tween 20 buffer and the solution then shaken off. The slide was next treated with FITC-labeled streptavidin at 10 μg/ml in the same buffer system for 30 min. After this time the streptavidin-buffer system was rinsed off with fresh PBS/0.05% Tween 20 buffer and then the slide was finally agitated in distilled water for about 1/2 h. After drying, the slide was examined by fluorescence microscopy (see Fig. 2 and Fig. 3).

4. substrate preparation

Before attachment of reactive groups it is preferred to clean the substrate which is, in a preferred embodiment, a glass substrate such as a microscope slide or cover slip. A roughened surface will be useable but a plastic or other solid substrate is also appropriate. According to one embodiment the slide is soaked in an alkaline bath consisting of, e.g., 1 liter of 95% ethanol with 120 ml of water and 120 grams of sodium hydroxide for 12 hours. The slides are washed with a

buffer and under running water, allowed to air dry, and rinsed with a solution of 95% ethanol.

The slides are then aminated with, e.g., aminopropyltriethoxysilane for the purpose of attaching amino groups to the glass surface on linker molecules, although other omega functionalized silanes could also be used for this purpose. In one embodiment 0.1% aminopropyltriethoxysilane is utilized, although solutions with concentrations from 10^{-7}% to 10% may be used, with about 10^{-3}% to 2% preferred. A 0.1% mixture is prepared by adding to 100 ml of a 95% ethanol/5% water mixture, 100 microliters (μl) of aminopropyltriethoxysilane. The mixture is agitated at about ambient temperature on a rotary shaker for an appropriate amount of time, e.g., about 5 minutes. 500 μl of this mixture is then applied to the surface of one side of each cleaned slide. After 4 minutes or more, the slides are decanted of this solution and thoroughly rinsed three times or more by dipping in 100% ethanol.

After the slides dry, they are heated in a 110-120°C vacuum oven for about 20 minutes, and then allowed to cure at room temperature for about 12 hours in an argon environment. The slides are then dipped into DMF (dimethylformamide) solution, followed by a thorough washing with methylene chloride.

5. linker attachment, blocking of free sites

The aminated surface of the slide is then exposed to about 500 μl of, for example, a 30 millimolar (mM) solution of NVOC-nucleotide- NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-nucleotide to each of the amino groups. See, e.g., SIGMA Chemical Company for various nucleotide derivatives. The surface is washed with, for example, DMF, methylene chloride, and ethanol.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-nucleotide attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this

residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

6. synthesis of eight trimers of C and T

Fig. 4 illustrates a possible synthesis of the eight trimers of the two-monomer set: cytosine and thymine (represented by C and T, respectively). A glass slide bearing silane groups terminating in 6-nitroveratryloxycarboxamide (NVOC-NH) residues is prepared as a substrate. Active esters (pentafluorophenyl, OBt, etc.) of cytosine and thymine protected at the 5' hydroxyl group with NVOC are prepared as reagents. While not pertinent to this example, if side chain protecting groups are required for the monomer set, these must not be photoreactive at the wavelength of light used to protect the primary chain.

For a monomer set of size n, n x ℓ cycles are required to synthesize all possible sequences of length ℓ. A cycle consists of:

1. Irradiation through an appropriate mask to expose the 5'-OH groups at the sites where the next residue is to be added, with appropriate washes to remove the by-products of the deprotection.
2. Addition of a single activated and protected (with the same photochemically-removable group) monomer, which will react only at the sites addressed in step 1, with appropriate washes to remove the excess reagent from the surface.

The above cycle is repeated for each member of the monomer set until each location on the surface has been extended by one residue in one embodiment. In other embodiments, several residues are sequentially added at one location before moving on to the next location. Cycle times will generally be limited by the coupling reaction rate, now as short as about 10 min in automated oligonucleotide synthesizers. This step is optionally followed by addition of

a protecting group to stabilize the array for later testing. For some types of polymers (e.g., peptides), a final deprotection of the entire surface (removal of photoprotective side chain groups) may be required.

More particularly, as shown in Fig. 4A, the glass 20 is provided with regions 22, 24, 26, 28, 30, 32, 34, and 36. Regions 30, 32, 34, and 36 are masked, indicated by the hatched regions, as shown in Fig. 4B and the glass is irradiated by the bright regions 22, 24, 26, and 28, and exposed to a reagent containing a photosensitive blocked C (e.g., cytosine derivative), with the resulting structure shown in Fig. 4C. The substrate is carefully washed and the reactants removed. Thereafter, regions 22, 24, 26, and 28 are masked, as indicated by the hatched region, the glass is irradiated (as shown in Fig. 4D), as indicated by the bright regions, at 30, 32, 34, and 36, and exposed to a photosensitive blocked reagent containing T (e.g., thymine derivative), with the resulting structure shown in Fig. 4E. The process proceeds, consecutively masking and exposing the sections as shown until the structure shown in Fig. 4M is obtained. The glass is irradiated and the terminal groups are, optionally, capped by acetylation. As shown, all possible trimers of cytosine/thymine are obtained.

In this example, no side chain protective group removal is necessary, as might be common in modified nucleotides. If it is desired, side chain deprotection may be accomplished by treatment with ethanedithiol and trifluoro-acetic acid.

In general, the number of steps needed to obtain a particular polymer chain is defined by:

$$n \times \ell \quad (1)$$

where:

n = the number of monomers in the basis set of monomers, and

ℓ = the number of monomer units in a polymer chain.

Conversely, the synthesized number of sequences of length ℓ will be:

$$n^{\ell}. \quad (2)$$

Of course, greater diversity is obtained by using masking strategies which will also include the synthesis of polymers having a length of less than ℓ. If, in the extreme case, all polymers having a length less than or equal to ℓ are synthesized, the number of polymers synthesized will be:

$$n^{\ell} + n^{\ell-1} + \ldots + n^{1}. \quad (3)$$

The maximum number of lithographic steps needed will generally be n for each "layer" of monomers, i.e., the total number of masks (and, therefore, the number of lithographic steps) needed will be n x ℓ. The size of the transparent mask regions will vary in accordance with the area of the substrate available for synthesis and the number of sequences to be formed. In general, the size of the synthesis areas will be:

size of synthesis areas = (A)/(S)

where:

A is the total area available for synthesis; and
S is the number of sequences desired in the area.

It will be appreciated by those of skill in the art that the above method could readily be used to simultaneously produce thousands or millions of oligomers on a substrate using the photolithographic techniques disclosed herein. Consequently, the method results in the ability to practically test large numbers of, for example, di, tri, tetra, penta, hexa, hepta, octa, nona, deca, even dodecanucleotides, or larger polynucleotides.

The above example has illustrated the method by way of a manual example. It will of course be appreciated that automated or semi-automated methods could be used. The substrate would be mounted in a flow cell for automated addition and removal of reagents, to minimize the volume of reagents needed, and to more carefully control reaction conditions. Successive masks will be applicable manually or automatically. See, e.g., PCT publication no. WO90/15070 and U.S.S.N. 07/624,120.

7. labeling of target

The target oligonucleotide can be labeled using standard procedures referred to above. As discussed, for certain situations, a reagent which recognizes interaction, e.g., ethidium bromide, may be provided in the detection step. Alternatively, fluorescence labeling techniques may be applied, see, e.g., Smith, et al. (1986) Nature, 321: 674-679; and Prober, et al. (1987) Science, 238:336-341. The techniques described therein will be followed with minimal modifications as appropriate for the label selected.

8. dimers of A, C, G, and T

The described technique may be applied, with photosensitive blocked nucleotides corresponding to adenine, cytosine, guanine, and thymine, to make combinations of polynucleotides consisting of each of the four different nucleotides. All 16 possible dimers would be made using a minor modification of the described method.

9. 10-mers of A, C, G, and T

The described technique for making dimers of A, C, G, and T may be further extended to make longer oligonucleotides. The automated system described, e.g., in PCT publication no. WO90/15070, and U.S.S.N. 07/624,120, can be adapted to make all possible 10-mers composed of the 4 nucleotides A, C, G, and T. The photosensitive, blocked nucleotide analogues have been described above, and would be readily adaptable to longer oligonucleotides.

10. specific recognition hybridization to 10-mers

The described hybridization conditions are directly applicable to the sequence specific recognition reagents attached to the substrate, produced as described immediately above. The 10-mers have an inherent property of hybridizing to a complementary sequence. For optimum discrimination between full matching and some mismatch, the conditions of hybridization should be carefully selected, as described above. Careful control of the conditions, and titration of parameters

should be performed to determine the optimum collective conditions.

11. hybridization

Hybridization conditions are described in detail, e.g., in Hames and Higgins (1985) Nucleic Acid Hybridisation: A Practical Approach; and the considerations for selecting particular conditions are described, e.g., in Wetmur and Davidson, (1988) J. Mol. Biol. 31:349-370, and Wood et al. (1985) Proc. Natl. Acad. Sci. USA 82:1585-1588. As described above, conditions are desired which can distinguish matching along the entire length of the probe from where there is one or more mismatched bases. The length of incubation and conditions will be similar, in many respects, to the hybridization conditions used in Southern blot transfers. Typically, the GC bias may be minimized by the introduction of appropriate concentrations of the alkylammonium buffers, as described above.

Titration of the temperature and other parameters is desired to determine the optimum conditions for specificity and distinguishability of absolutely matched hybridization from mismatched hybridization.

A fluorescently labeled target or set of targets are generated, as described in Prober, et al. (1987) Science 238:336-341, or Smith, et al. (1986) Nature 321:674-679. Preferably, the target or targets are of the same length as, or slightly longer, than the oligonucleotide probes attached to the substrate and they will have known sequences. Thus, only a few of the probes hybridize perfectly with the target, and which particular ones did would be known.

The substrate and probes are incubated under appropriate conditions for a sufficient period of time to allow hybridization to completion. The time is measured to determine when the probe-target hybridizations have reached completion. A salt buffer which minimizes GC bias is preferred, incorporating, e.g., buffer, such as tetramethyl ammonium or tetraethyl ammonium ion at between about 2.4 and 3.0 M. See Wood, et al. (1985) Proc. Nat'l Acad. Sci. USA 82:1585-1588.

This time is typically at least about 30 min, and may be as long as about 1-5 days. Typically very long matches will hybridize more quickly, very short matches will hybridize less quickly, depending upon relative target and probe concentrations. The hybridization will be performed under conditions where the reagents are stable for that time duration.

Upon maximal hybridization, the conditions for washing are titrated. Three parameters initially titrated are time, temperature, and cation concentration of the wash step. The matrix is scanned at various times to determine the conditions at which the distinguishability between true perfect hybrid and mismatched hybrid is optimized. These conditions will be preferred in the sequencing embodiments.

12. positional detection of specific interaction

As indicated above, the detection of specific interactions may be performed by detecting the positions where the labeled target sequences are attached. Where the label is a fluorescent label, the apparatus described, e.g., PCT publication no. WO90/15070; and U.S.S.N. 07/624,120, may be advantageously applied. In particular, the synthetic processes described above will result in a matrix pattern of specific sequences attached to the substrate, and a known pattern of interactions can be converted to corresponding sequences.

In an alternative embodiment, a separate reagent which differentially interacts with the probe and interacted probe/targets can indicate where interaction occurs or does not occur. A single-strand specific reagent will indicate where no interaction has taken place, while a double-strand specific reagent will indicate where interaction has taken place. An intercalating dye, e.g., ethidium bromide, may be used to indicate the positions of specific interaction.

13. analysis

Conversion of the positional data into sequence specificity will provide the set of subsequences whose analysis by overlap segments, may be performed, as described above.

Analysis is provided by the methodology described above, or using, e.g., software available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia (Yugoslav group). See, also, Macevicz, PCT publication no. WO 90/04652, which is hereby incorporated herein by reference.

The description of the preparation of short peptides on a substrate incorporates by reference sections in U.S.S.N. 07/492,462 (VLSIPS CIP), and described below.

POLYNUCLEOTIDE FINGERPRINTING

The above section on generation of reagents for sequencing provides specific reagents useful for fingerprinting applications. Fingerprinting embodiments may be applied towards polynucleotide fingerprinting, cell and tissue classification, cell and tissue temporal development stage classification, diagnostic tests, forensic uses for individual identification, classification of organisms, and genetic screening of individuals. Mapping applications are also described below.

Polynucleotide fingerprinting may use reagents similar to those described above for probing a sequence for the presence of specific subsequences found therein. Typically, the subsequences used for fingerprinting will be longer than the sequences used in oligonucleotide sequencing. In particular, specific long segments may be used to determine the similarity of different samples of nucleic acids. They may also be used to fingerprint whether specific combinations of information are provided therein. Particular probe sequences are selected and attached in a positional manner to a substrate. The means for attachment may be either using a caged biotin method described, e.g., in U.S.S.N. 07/612,671 (caged biotin CIP), or by another method using targeting molecules. In one embodiment, an unnatural nucleotide or similar complementary binding molecule may be attached to the fingerprinting probe and the probe thereby directed towards complementary sequences on a VLSIPS substrate. Typically, unnatural nucleotides would be preferred, e.g., unnatural

optical isomers, which would not interfere with natural nucleotide interactions.

Having produced a substrate with particular fingerprint probes attached thereto at positionally defined regions, the substrate may be used in a manner quite similar to the sequencing embodiment to provide information as to whether the fingerprint probes are detecting the corresponding sequence in a target sequence. This will often provide information similar to a Southern blot hybridization.

Temporal Development

Developmental RNA expression patterns

The present fingerprinting invention also allows cell classification by identification of developmental RNA expression patterns. For example, a lymphocyte stem cell expresses a particular combination of RNA species. As the lymphocyte develops through a program developmental scheme, at various stages it expresses particular RNA species which are diagnostic of particular stages in development. Again, the fingerprinting methodology allows for the definition of specific structural features which are diagnostic of developmental or functional features which will allow classification of cells into temporal developmental classes. Cells, products of those cells, or lysates of those cells will be assayed to determine the developmental stage of the source cells. In this manner, once a developmental stage is defined, specific synchronized populations of cells will be selected out of another population. These synchronized populations may be very important in determining the biological mechanisms of development.

The present invention also allows for fingerprinting of the mRNA population of a cell. In this fashion, the mRNA population, which should be a good determinant of developmental stage, will be correlated with other structural features of the cell. In this manner, cells at specific developmental stages will be characterized by the intracellular environment, as well as the extracellular environment.

Diagnostic Tests

The present invention also provides the ability to perform diagnostic tests. Diagnostic tests typically are based upon a fingerprint type assay, which tests for the presence of specific diagnostic polynucleotides. Thus, the present invention provides means for viral strain identification, bacterial strain identification, and other diagnostic tests using positionally defined specific oligonucleotide reagents.

Viral Identification

The present invention provides reagents and methodology for identifying viral strains. The viral genome may be probed for specific sequences which are characteristic of particular viral strains. Specific hybridization patterns on an VLSIPS oligonucleotide substrate can identify the presence of particular viral genomes.

Bacterial Identification

Similar techniques will be applicable to identifying a bacterial source. This may be useful in diagnosing bacterial infections, or in classifying sources of particular bacterial species. For example, the bacterial assay may be useful in determining the natural range of survivability of particular strains of bacteria across regions of the country or in different ecological niches.

Other Microbiological Identifications

The present invention provides means for diagnosis of other microbiological and other species, e.g., protozoal species and parasitic species in a biological sample, but also provides the means for assaying a combination of different infections. For example, a biological specimen may be assayed for the presence of any or all of these microbiological species. In human diagnostic uses, typical samples will be blood, sputum, stool, urine, or other samples.

Individual Identification

The present invention provides the ability to fingerprint and identify a genetic individual. This individual may be a bacterial or lower microorganism, as described above in diagnostic tests, or of a plant or animal. An individual may be identified genetically, as described.

Genetic fingerprinting has been utilized in comparing different related species in Southern hybridization blots. Genetic fingerprinting has also been used in forensic studies, see, e.g., Morris et al. (1989) J. Forensic Science 34: 1311-1317, and references cited therein. As described above, an individual may be identified genetically by a sufficiently large number of probes. The likelihood that another individual would have an identical pattern over a sufficiently large number of probes may be statistically negligible. However, it is often quite important that a large number of probes be used where the statistical probability of matching is desired to be particularly low. In fact, the probes will optimally be selected for having high heterogeneity among the population. In addition, the fingerprint method may make use of the pattern of homologies indicated by a series of more and more stringent washes. Then, each position has both a sequence specificity and a homology measurement, the combination of which greatly increases the number of dimensions and the statistical likelihood of a perfect pattern match with another genetic individual.

Genetic Screening

1. test alleles with markers

The present invention provides for the ability to screen for genetic variations of individuals. For example, a number of genetic diseases are linked with specific alleles. See, e.g., Scriber, C. et al. (eds.) (1989) The Metabolic Bases of Inherited Disease, McGraw-Hill, New York. In one embodiment, cystic fibrosis has been correlated with a specific gene, see, Gregory et al. (1990) Nature 347: 382-386. A number of alleles are correlated with specific genetic deficiencies. See, e.g., McKusick, V. (1990) Genetic Inheritance in Man:

Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes, Johns Hopkins University Press, Baltimore; Ott, J. (1985) Analysis of Human Genetic Linkage, Johns Hopkins University Press, Baltimore; Track, R. et al. (1989) Banbury Report 32: DNA Technology and Forensic Science, Cold Spring Harbor Press, New York; each of which is hereby incorporated herein by reference.

2. Amniocentesis

Typically, amniocentesis is used to determine whether chromosome translocations have occurred. The mapping procedure may provide the means for determining whether these translocations have occurred, and for detecting particular alleles of various markers.

MAPPING

Positionally Located Clones

The present invention allows for the positional location of specific clones useful for mapping. For example, caged biotin may be used for specifically positioning a probe to a location on a matrix pattern.

In addition, the specific probes may be positionally directed to specific locations on a substrate by targeting. For example, polypeptide specific recognition reagents may be attached to oligonucleotide sequences which can be complementarily targeted, by hybridization, to specific locations on a VLSIPS substrate. Hybridization conditions, as applied for oligonucleotide probes, will be used to target the reagents to locations on a substrate having complementary oligonucleotides synthesized thereon. In another embodiment, oligonucleotide probes may be attached to specific polypeptide targeting reagents such as an antigen or antibody. These reagents can be directed towards a complementary antigen or antibody already attached to a VLSIPS substrate.

In another embodiment, an unnatural nucleotide which does not interfere with natural nucleotide complementary hybridization may be used to target oligonucleotides to

particular positions on a substrate. Unnatural optical isomers of natural nucleotides should be ideal candidates.

In this way, short probes may be used to determine the mapping of long targets or long targets may be used to map the position of shorter probes. See, e.g., Craig et al. 1990 Nuc. Acids Res. 18: 2653-2660.

Positionally Defined Clones

Positionally defined clones may be transferred to a new substrate by either physical transfer or by synthetic means. Synthetic means may involve either a production of the probe on the substrate using the VLSIPS synthetic methods, or may involve the attachment of a targeting sequence made by VLSIPS synthetic methods which will target that positionally defined clone to a position on a new substrate. Both methods will provide a substrate having a number of positionally defined probes useful in mapping.

CONCLUSION

The present inventions provide greatly improved methods and apparatus for synthesis of polymers on substrates. It is to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of skill in the art upon reviewing the above description. By way of example, the invention has been described primarily with reference to the use of photoremovable protective groups, but it will be readily recognized by those of skill in the art that sources of radiation other than light could also be used. For example, in some embodiments it may be desirable to use protective groups which are sensitive to electron beam irradiation, x-ray irradiation, in combination with electron beam lithograph, or x-ray lithography techniques. Alternatively, the group could be removed by exposure to an electric current. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

All publications and patent applications referred to herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually incorporated by reference. The present invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

WHAT IS CLAIMED IS:

1. A composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid substrate, which reagents are capable of specifically binding to a predetermined subunit sequence of a preselected multi-subunit length having at least five subunits, said reagents representing substantially all possible sequences of said preselected length.

2. A composition of Claim 1, wherein said subunit sequence is a polynucleotide.

3. A composition of Claim 1, wherein said specific reagent is an polynucleotide of at least about eight nucleotides.

4. A composition of Claim 1, wherein said specific reagents are all attached to a single solid substrate.

5. A composition of Claim 1, wherein said reagents comprise about 3000 different sequences.

6. A composition of Claim 1, wherein said reagents represents at least about 25% of the possible subsequences of said preselected length.

7. A composition of Claim 1, wherein said reagents are localized in regions of the substrate having a density of at least 25 regions per square centimeter.

8. A composition of Claim 4, wherein said substrate has a surface area of less than about 4 square centimeters.

9. A method of analyzing a sequence of a polynucleotide, said method comprising the step of:

a) exposing said polynucleotide to a composition of Claim 1.

10. A method of identifying or comparing a target sequence with a reference, said method comprising the step of:

a) exposing said target sequence to a composition of Claim 1;

b) determining the pattern of positions of said reagents which specifically interact with said target sequence; and

c) comparing said pattern with the pattern exhibited by said reference when exposed to said composition.

11. A method for sequencing a segment of a polynucleotide comprising the steps of:

a) combining:

i) a substrate comprising a plurality of chemically synthesized and positionally distinguishable oligonucleotides capable of recognizing defined oligonucleotide sequences; and

ii) a target polynucleotide; thereby forming high fidelity matched duplex structures of complementary subsequences of known sequence; and

b) determining which of said reagents have specifically interacted with subsequences in said target polynucleotide.

12. A method of Claim 11, wherein said segment is substantially the entire length of said polynucleotide.

13. A method for sequencing a polymer, said method comprising the steps of:

a) preparing a plurality of reagents which each specifically bind to a subsequence of preselected length;

b) positionally attaching each of said reagents to one or more solid phase substrates, thereby producing substrates of positionally definable sequence specific probes;

c) combining said substrates with a target polymer whose sequence is to be determined; and

d) determining which of said reagents have specifically interacted with subsequences in said target polymer.

14. A method of Claim 13, wherein said substrates are beads.

15. A method of Claim 13, wherein said plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target.

16. A method of Claim 13, wherein said solid phase substrates are a single substrate having attached thereto reagents recognizing substantially all possible subsequences of preselected length found in said target.

17. A method of Claim 13, further comprising the step of analyzing a plurality of said recognized subsequences to assemble a sequence of said target polymer.

18. A method of Claim 14, wherein at least some of said plurality of substrates have one subsequence specific reagent attached thereto, and said substrates are coded to indicate the specificity of said reagent.

19. A method of using a fluorescent nucleotide tc detect interactions with oligonucleotide probes of known sequence, said method comprising:

a) attaching said nucleotide to a target unknown polynucleotide sequence, and

b) exposing said target polynucleotide sequence to a collection of positionally defined oligonucleotide probes of known

sequences to determine the sequences of said probes which interact with said target.

20. A method of Claim 19, further comprising the step of:

a) collating said known sequences to determine the overlaps of said known sequences to determine the sequence of said target sequence.

21. A method of mapping a plurality of sequences relative to one another, said method comprising:

a) preparing a substrate having a plurality of positionally attached sequence specific probes are attached;

b) exposing each of said sequences to said substrate, thereby determining the patterns of interaction between said sequence specific probes and said sequences; and

c) determining the relative locations of said sequence specific probe interactions on said sequences to determine the overlaps and order of said sequences.

22. A method of Claim 21, wherein said sequence specific probes are oligonucleotides.

23. A method of Claim 21, wherein said sequences are nucleic acid sequences.

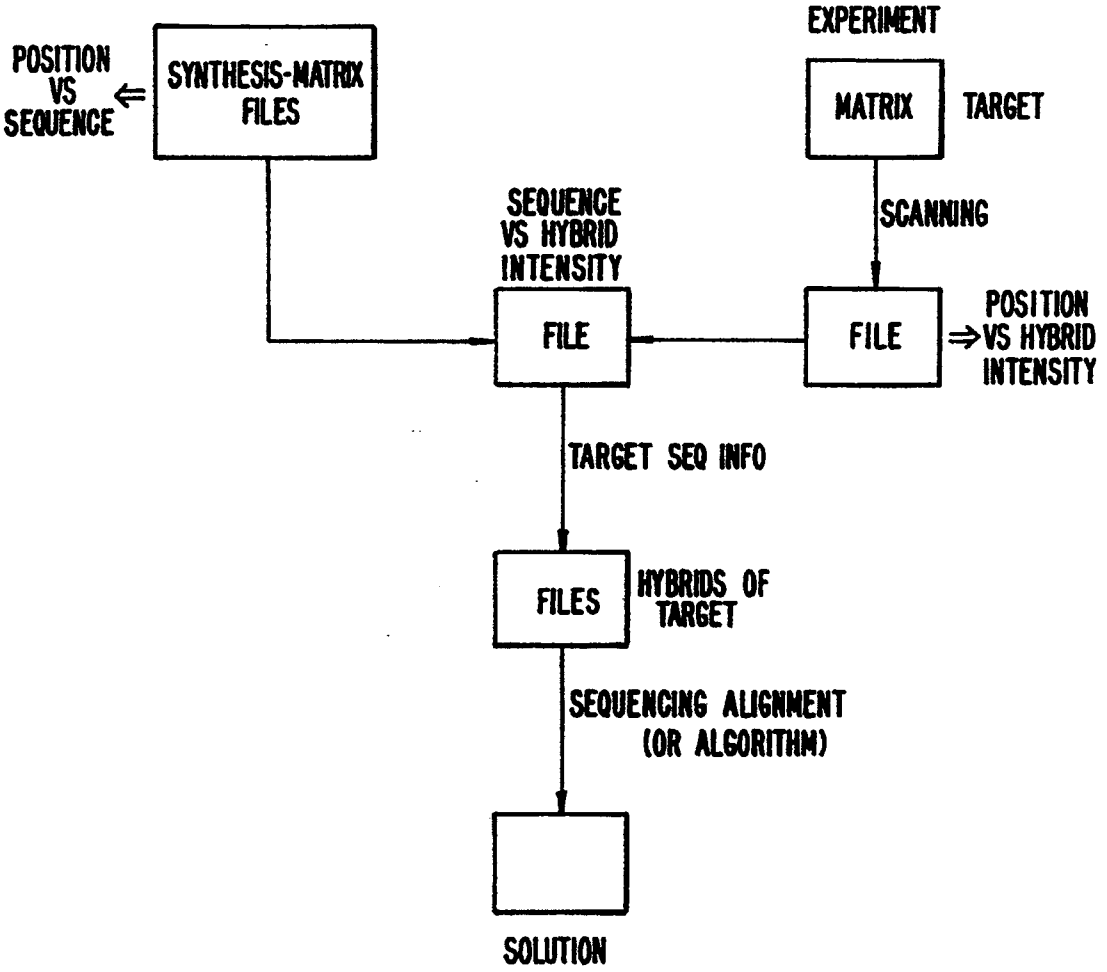

EXPERIMENT
POSITION VS SEQUENCE
SYNTHESIS-MATRIX FILES
MATRIX
TARGET
SCANNING
SEQUENCE VS HYBRID INTENSITY
FILE
FILE
POSITION VS HYBRID INTENSITY
TARGET SEQ INFO
FILES
HYBRIDS OF TARGET
SEQUENCING ALIGNMENT (OR ALGORITHM)
SOLUTION

FIG. 1.

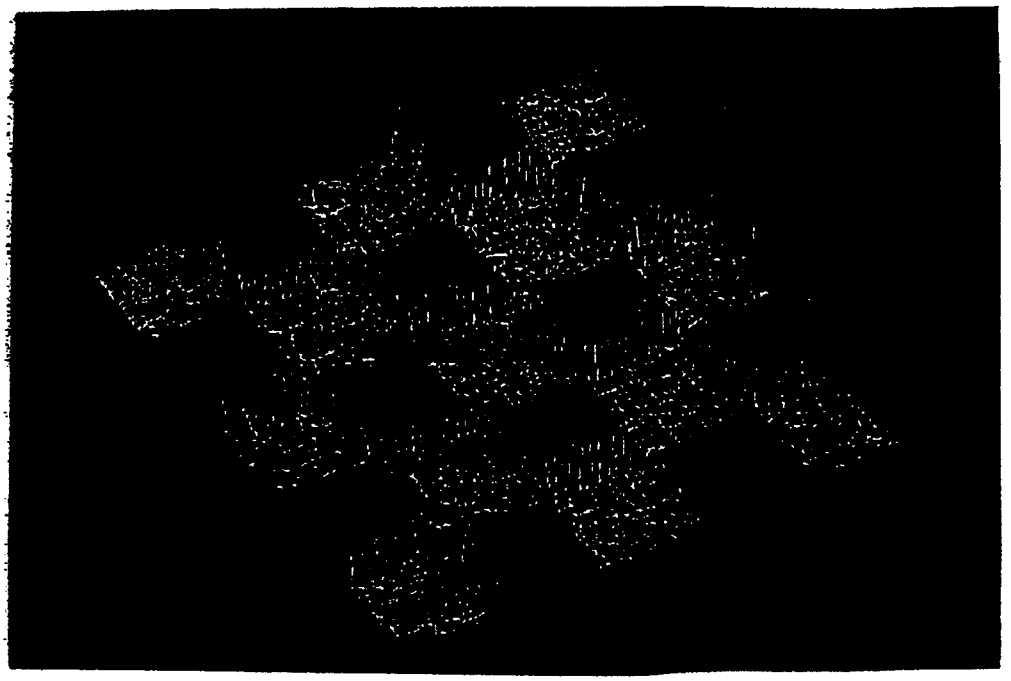


FIG. 2.

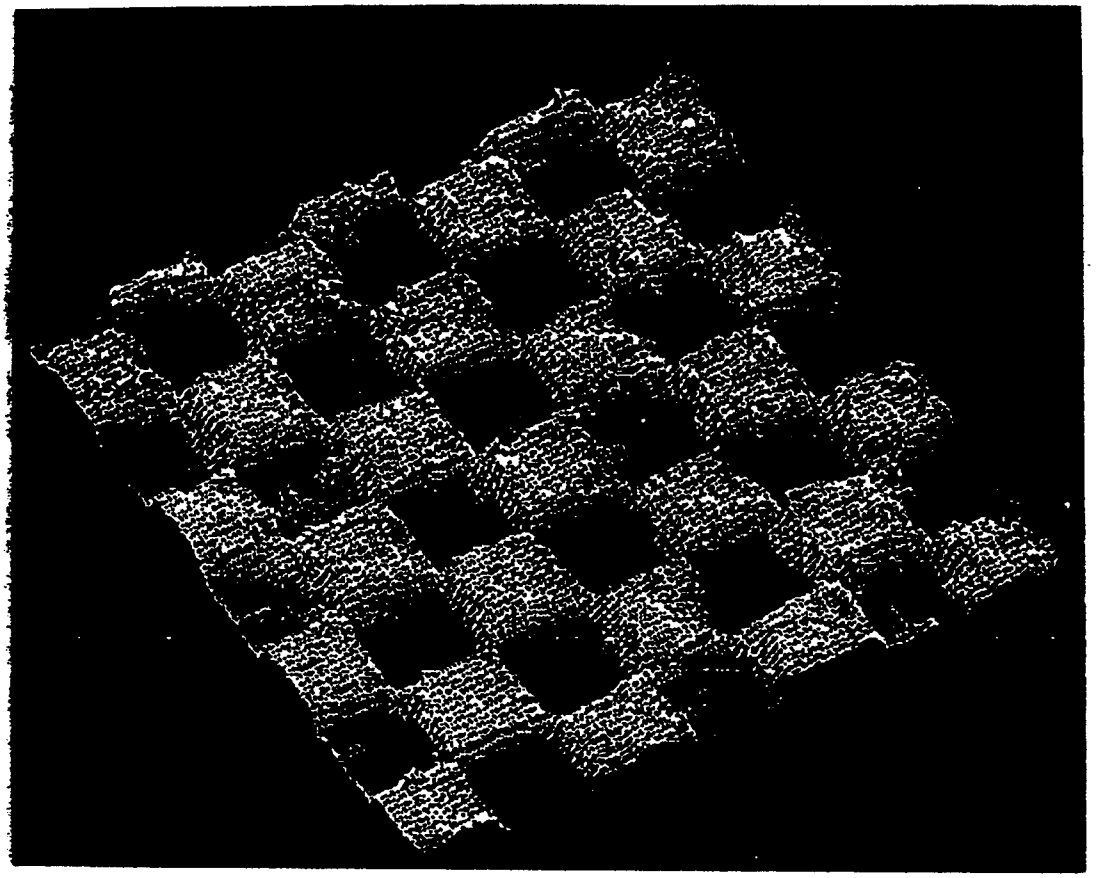


FIG. 3.

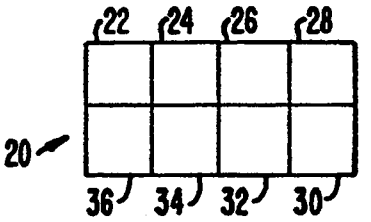


FIG. 4A.

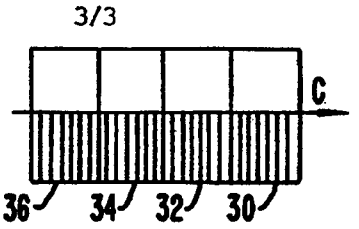


FIG. 4B.

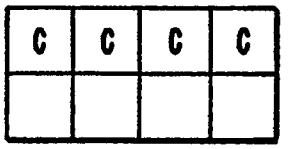


FIG. 4C.

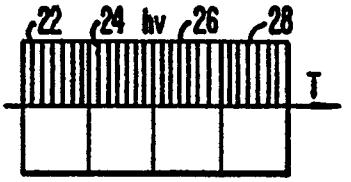


FIG. 4D.

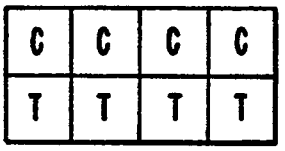


FIG. 4E.

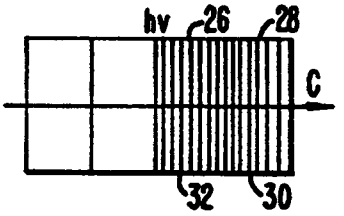


FIG. 4F.

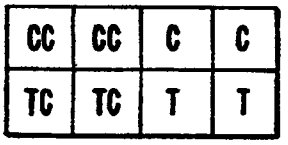


FIG. 4G.

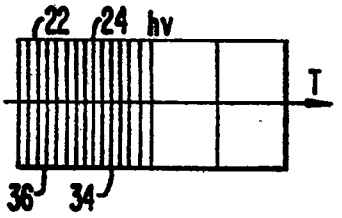


FIG. 4H.

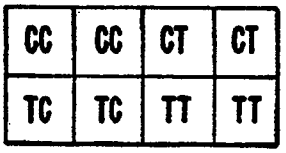


FIG. 4I.

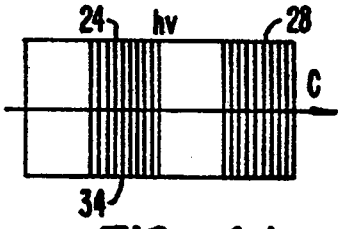


FIG. 4J.

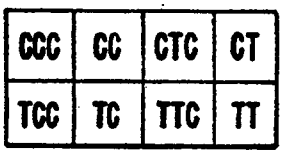


FIG. 4K.

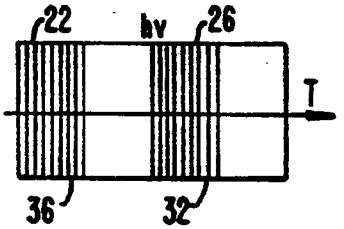


FIG. 4L.

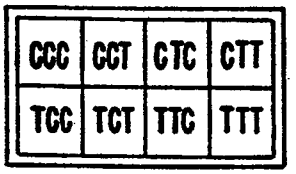


FIG. 4M.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/09226

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C12Q 1/68; G01N 33/566, 33/48; C07H 15/12
US CL : 435/6; 436/501, 94; 536/26,27,28,29; 935/77,78

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	435/6; 436/501, 94; 536/26,27,28,29; 935/77,78

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[5]

Please See Attached Sheet.

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
X/Y	Doklady Akademii Nauk SSSR, Vol. 303, No. 6, issued December 1988, Yu. P. Lysov et al., "A New Method for Determining the DNA Nucleotide Sequence by Hybridization with Oligonucleotides", pages 436-438, see page 437, paragraph 3 and page 436, paragraph 2 (translation, Plenum Publishing Corporation, 1989).	1-13, 15-17, 18-23/14, 18

* Special categories of cited documents:[15]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
10 MARCH 1992	17 MAR 1992
International Searching Authority[1]	Signature of Authorized Officer[20]
ISA/US	Stephanie W. Zitomer, Ph.D.

FURTHER INFORMATION CONTINUED FROM PREVIOUS SHEETS
(Not for publication)

II. FIELDS SEARCHED
Other Documents Searched:

Dialog Onesearch: Biosis, Derwent Biotech. Abstracts, World Patent Index, Jap. Tech. Abstracts, CA; APS; search terms: (oligonucleotide# or polynucleotide#), ((immobiliz? or matrix), (sequencing or sequence()determination)